ACCURATE LINEAR PARAMETER
ESTIMATION WITH NOISY INPUTS

Inventor(s):

Deniz Erdogmus

Jose C. Principe

Yadunandana N. Rao

University of Florida

Disclosure No. 11173

Express Mailing Label No. EV 347797909 US

ACCURATE LINEAR PARAMETER
ESTIMATION WITH NOISY INPUTS

CROSS REFERENCE TO RELATED APPLICATIONS

[0001]   This application claims the benefit of U.S. Provisional Application No. 60/459,285, filed in the United States Patent and Trademark Office on March 31, 2003, the entirety of which is incorporated herein by reference.

BACKGROUND

Field of the Invention

[0002]   The present invention relates to the field of filter adaptation and system identification in the presence of noise.

Description of the Related Art

[0003]   System identification refers to the construction of mathematical models of a dynamic system based upon measured data.   One type of mathematical model used to emulate the behavior of physical plants is the linear Autoregressive Moving Average (ARMA) model.   In developing such a model, typically parameters of the model are adjusted until the output of the model coincides with that of the actual system output.   The accuracy of a model can be evaluated using conventional Mean Squared Error (MSE) techniques to compare the actual system output with the predicted output of the mathematical model.

[0004]   System identification is an important aspect of designing controllers for physical plants.   Accurate system identification facilitates the design of robust controllers.   System identification, however, can be imprecise when sensors of the physical plant being modeled collect noise in addition to data.

[0005]   One conventional method of dealing with noise has been to condition a received signal in the hopes of minimizing or removing noise.   Because noise and signal bands overlap in most cases, signal conditioning or filtering, at best, presents a compromise in that the noise cannot be removed from the signal bands.

[0006]   Moreover, conventional MSE-based techniques are not useful indicators of model accuracy when data has been corrupted with additive white noise or noise which is similar to, or

can be modeled as white noise. It has been widely acknowledged that MSE is optimal for linear filter estimation when there are no noisy perturbations on the data. For many real-world applications, however, the "noise-free" assumption is easily violated and using MSE-based methods for parameter estimation can result in severe parameter bias.

[0007] What is needed is a technique for estimating model parameters for a physical plant which provides acceptable results in the presence of noise.

SUMMARY OF THE INVENTION

[0008] The inventive arrangements disclosed herein provide a method, system, and apparatus for linear model parameter estimation in the presence of white noise, or noise that can be so approximated. Also provided is a novel criterion for performing filter adaptation. It should be appreciated that the more the actual noise in a system resembles white noise, the more accurate the results obtained from the inventive arrangements disclosed herein. The present invention further can reduce the number of parameters needed to model an unknown system when compared with mean squared error-based techniques.

[0009] One embodiment of the present invention can include a method of building a model for a physical plant in the presence of noise. The method can include initializing a model of the physical plant, wherein the model is characterized by a parameter vector, estimating an output of the model, and computing a composite cost comprising a weighted average of a squared error between the estimated output from the model and an actual output of the physical plant, and a squared derivative of the error. The method further can include determining a step size and an update direction. The model of the physical plant can be updated. Notably, the updating step can be dependent upon the step size.

[0010] Another embodiment of the present invention can include initializing a model of a physical plant and an inverse Hessian matrix, wherein the model is characterized by a parameter vector. The method also can include determining a Kalman gain, estimating the output of the model, and computing a composite cost comprising a weighted average of an error vector between the estimated output from the model and an actual output of the physical plant, and a derivative of the error. The method further can include updating the model of the physical plant and updating the inverse Hessian matrix.

[0011] Other embodiments of the present invention can include a system having means for performing the various steps disclosed herein as well as a machine readable storage for causing a machine to perform the steps described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] There are shown in the drawings, embodiments which are presently preferred, it being understood, however, that the invention is not limited to the precise arrangements and instrumentalities shown.

[0013] FIG. 1 is a schematic diagram illustrating a system in which embodiments of the present invention can be used.

[0014] FIG. 2 is a flow chart illustrating a method of performing linear parameter estimation for a physical plant in accordance with one embodiment of the present invention.

[0015] FIG. 3 is a flow chart illustrating a method of performing linear parameter estimation for a physical plant in accordance with another embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

[0016] FIG. 1 is a schematic diagram illustrating a system 100 in which embodiments of the present invention can be used. As shown, the system 100 can include a physical plant 105 and a model 110. The physical plant 105 can be any of a variety of physical machines, whether simple or complex, for which a linear model can be constructed to estimate the behavior of the physical plant 105. For example, the physical plant 105 can be a combustion engine, an assembly line, a manufacturing process, a biomedical process, or the like. The model 110 can be a linear, software-based model of the physical plant 105 that executes within a suitable information processing system.

[0017] Referring to FIG. 1, (x_k, d_k) can denote the actual input and output of the physical plant 105. Measurement errors and system disturbances can be modeled as uncorrelated additive white noise sequences u_k and v_k having unknown variances that appear at the output and input of the physical plant 105 respectively.

[0018] System identification can be performed to build the model 110 of the physical plant 105. Given the noisy data pair $(\hat{x}_k, \hat{d}_k)$, where $\hat{x}_k \in \Re^N = x_k + v_k$ and $\hat{d}_k \in \Re^1 = d_k + u_k$, a parameter vector $w \in \Re^M$ can be determined that suitably describes the physical plant 105.

Without loss of generality, the length of $\mathbf{w}$ can be assumed to be at least N, the number of parameters in the actual system, i.e. $M \geq N$, where M denotes the length of the parameter vector to be estimated. M also can be referred to as the model-length or the model-order. Since $d_k = \mathbf{x}_k^T \mathbf{w}_T$, the error can be calculated as $\hat{e}_k = \mathbf{x}_k^T (\mathbf{w}_T - \mathbf{w}) + u_k - \mathbf{v}_k^T \mathbf{w}$.

[0019]   Defining a vector $\boldsymbol{\varepsilon} = \mathbf{w}_T - \mathbf{w}$, the error autocorrelation at some arbitrary lag L can be determined using equation (1).

$$(1) \qquad \rho_{\hat{e}}(L) = \boldsymbol{\varepsilon}^T E[\mathbf{x}_k \mathbf{x}_{k-L}^T] \boldsymbol{\varepsilon} + \mathbf{w}^T E[\mathbf{v}_k \mathbf{v}_{k-L}^T] \mathbf{w}$$

The error autocorrelation is a measure of the time structure of the error signal. The lag is a scalar value which can be chosen to measure the similarity between the error at a current time instant t and the error at time t-L. If the chosen lag, $L \geq M$ then $E[\mathbf{v}_k \mathbf{v}_{k-L}^T] = \mathbf{0}$. In that case, the error autocorrelation can be represented as equation (2).

$$(2) \qquad \rho_{\hat{e}}(L) = \boldsymbol{\varepsilon}^T E[\mathbf{x}_k \mathbf{x}_{k-L}^T] \boldsymbol{\varepsilon} = (\mathbf{w}_T - \mathbf{w})^T E[\mathbf{x}_k \mathbf{x}_{k-L}^T](\mathbf{w}_T - \mathbf{w})$$

[0020]   If the matrix $E[\mathbf{x}_k \mathbf{x}_{k-L}^T]$ is full rank, $\rho_{\hat{e}}(L) = 0$ only when $\mathbf{w} = \mathbf{w}_T$. Therefore, if the error autocorrelation at any lag $L \geq M$ is made to be zero, the estimated weight vector will be exactly equal to the true weight vector. In other words, the criterion tries to whiten the error signal for lags greater than or equal to the adaptive filter length, i.e., $\rho_{\hat{e}}(L) = 0$ for $L \geq M$. As such, the criterion can be referred to as the Error Whitening Criterion (EWC). In other words, if the error is partially whitened, the estimated model has captured the relevant information present in the input and output data. In general, a white signal carries no meaningful information. By making the error signal partially white, EWC extracts the essential information from the input/output data and captures that information within the model parameters. EWC can be represented as equation (3) below.

[0021]   Defining $\dot{\hat{e}}_k = (\hat{e}_k - \hat{e}_{k-L})$, equation (1) can be rewritten as

$$(3) \qquad J(\mathbf{w}) = E(\hat{e}_k^2) + \beta E(\dot{\hat{e}}_k^2)$$

where β is a constant. Setting $\beta = -0.5$ and restricting $L \geq M$, equation (3) can be reduced to the error autocorrelation $\rho_{\hat{e}}(L)$ given by equation (2). Accordingly, a weight vector $\mathbf{w}$ can be found that makes $J(\mathbf{w}) = 0$ with $\beta = -0.5$. In one embodiment, when $\beta = 0$, EWC reduces to the Mean Squared Error (MSE) cost function. The derivative of $\rho_{\hat{e}}(L)$ with respect to $\mathbf{w}$ can be

determined by $\partial \rho_{\hat{e}}(L)/\partial \mathbf{w} = -2[\mathbf{w}_T - \mathbf{w}]E[\mathbf{x}_k \mathbf{x}_{k-L}^T]$, and is zero when $(\mathbf{w}_T - \mathbf{w}) = \mathbf{0}$ and

$E[\mathbf{x}_k \mathbf{x}_{k-L}^T]$ is full rank. Thus, $\rho_{\hat{e}}(L) = 0$ and $\partial \rho_{\hat{e}}(L)/\partial \mathbf{w} = \mathbf{0}$ simultaneously when $\mathbf{w} = \mathbf{w}_T$.

[0022] Equation (4) defines a stochastic gradient update for online, local adaptation based upon EWC. This technique can be referred to as EWC-Least Mean Squares (LMS).

$$(4) \qquad \mathbf{w}_{k+1} = \mathbf{w}_k + \eta \, sign(\hat{e}_k^2 + \beta\hat{\hat{e}}_k^2)(\hat{e}_k \hat{\mathbf{x}}_k + \beta\hat{\hat{e}}_k \hat{\hat{\mathbf{x}}}_k)$$

[0023] Equation (4) includes the sign term that instantaneously changes the sign of the gradient. The expression $sign(\hat{e}_k^2 + \beta\hat{\hat{e}}_k^2)$ represents an update direction. The sign term accommodates the result that the error autocorrelation at arbitrary lags can take either positive or negative values. This means that the stationary point of equation (3) can be either a global minimum, maximum, or saddle point. Equation (4) converges under the conditions listed below.

[0024] In the noisy data case, the stochastic algorithm in equation (4) with $\beta = -0.5$ converges to the stationary point $\mathbf{w}_* = \mathbf{w}_T$ in the mean provided that the step size η is bound by the inequality specified in equation (5).

$$(5) \qquad 0 < \eta < \frac{2\left|E(\hat{e}_k^2 - 0.5\hat{\hat{e}}_k^2)\right|}{E\left\|\hat{e}_k \hat{\mathbf{x}}_k - 0.5\hat{\hat{e}}_k \hat{\hat{\mathbf{x}}}_k\right\|^2}$$

The conditions shown above are necessary for asymptotic convergence. Notably, the bound for the step-size in equation (5) can be computed in a practical and useful manner without the use of significant computing resources.

[0025] Further, with $\beta = -0.5$, the steady state ($\mathbf{w} = \mathbf{w}_*$) excess error autocorrelation at lag $L \geq M$, i.e., $|\rho_{\hat{e}}(L)|$ is always bound by,

$$(6) \qquad |\rho_{\hat{e}}(L)| \leq \frac{\eta}{2} E(\hat{e}_\infty^2)[Tr(\mathbf{R} + \mathbf{V})] + 2\eta[\sigma_u^2 + \|\mathbf{w}_\infty\|\|\mathbf{w}_*\|Tr(\mathbf{V})]$$

where $\mathbf{R} = E[\mathbf{x}_k \mathbf{x}_k^T]$, and $\mathbf{V} = E[\mathbf{v}_k \mathbf{v}_k^T]$ and $Tr(\bullet)$ denotes the matrix trace. The noise variances in the input and desired signals are represented by σ_v^2 and σ_u^2 respectively. As such, using the stochastic technique, the misadjustment can be arbitrarily minimized by having a time varying step-size that asymptotically decays to zero.

[0026] FIG. 2 is a flow chart illustrating a method 200 of performing linear parameter estimation for a physical plant in accordance with one embodiment of the present invention.

Method 200 illustrates the EWC-LMS technique described above and has a complexity of $O(M)$, where M is the number of parameters to be estimated. The method 200 can begin in a state where a physical plant is to be modeled with a linear model. As noted, the system or physical plant being modeled can include noise.

[0027] The method can begin in step 205 wherein model parameters are initialized. In particular, a weight vector characterizing the model denoted as $\mathbf{w}_k$ can be set equal to some initial value. In one embodiment, the weight vector can be initialized to 0. The step size η and β also can be initialized. While β can be set to any of a variety of different values, in one embodiment, β can be set equal to -0.5 or substantially equal to -0.5. Further, a lag L can be selected. As noted, the lag can be assigned a value that is greater than or equal to the number of parameters in the system including the physical plant.

[0028] In step 210, the output of the model can be computed. The output d_k can be estimated according to $d_k = \mathbf{x}_k^T \mathbf{w}_T$. In step 215, the error between the estimated output and the actual physical plant output can be determined. The current input vector to the physical plant and the output, which typically is a scalar, can be measured. Accordingly, the error e_k can be calculated according to $e_k = d_k - y_k$.

[0029] In step 220, the derivative of the input vector to the physical plant can be determined. In step 225, the derivative of the error can be calculated. Using the error and the error derivative, the composite cost can be determined. The composite cost can be calculated as a weighted average of the squared error between the estimated output of the model and the actual output of the physical plant, and the squared derivative of the error. Further the update direction also can be determined. In step 230, the model of the physical plant can be updated. That is, the model of the physical plant which is characterized by the weight vector $\mathbf{w}_k$ can be updated according to

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \eta sign(\hat{e}_k^2 + \beta\hat{\dot{e}}_k^2)(\hat{e}_k\hat{\mathbf{x}}_k + \beta\hat{\dot{e}}_k\hat{\dot{\mathbf{x}}}_k).$$

[0030] The method 200 can loop back to step 210 to repeat as necessary until a solution is determined or the method converges.

[0031] FIG. 3 is a flow chart illustrating a method 300 of performing linear parameter estimation for a physical plant in accordance with another embodiment of the present invention. The method 300 provides a fast Quasi-Newton type recursive technique for finding the stationary

point of the EWC. The complexity of this technique, referred to as the Recursive Error Whitening (REW) technique, is $O(M^2)$, where M is the number of parameters to be estimated. Though the REW technique has a higher complexity than the EWC-LMS technique described above, the REW technique is fast converging and functions independently of eigenspread and other issues associated with gradient methods.

[0032] The method 300 can begin in step 305 where the model parameters can be initialized. More particularly, $\mathbf{w}_k$ can be initialized to some beginning value. In one embodiment, $\mathbf{w}_k$ can be set equal to 0. Further, the inverse Hessian matrix $\mathbf{Z}_0^{-1}$ can be initialized to $c\mathbf{I}$, where c can be a large positive constant. In one embodiment, c can be set to a value between approximately 100-1,000. The matrix $\mathbf{I}$ denotes an identity matrix. The inverse Hessian matrix $\mathbf{Z}_0^{-1}$ is the second derivative of the criterion with respect to the paramters $\mathbf{w}$. As noted, while β can be set to any of a variety of different values, in one embodiment, β can be set equal to -0.5 or substantially equal to -0.5.

[0033] In step 310, matrices $\mathbf{B}$ and $\mathbf{D}$ can be defined. Matrix $\mathbf{B}$ can be defined as $[(2\beta\hat{\mathbf{x}}_k - \beta\hat{\mathbf{x}}_{k-L}) \quad \hat{\mathbf{x}}_k]$ and matrix $\mathbf{D}$ can be defined as $[\hat{\mathbf{x}}_k \quad (\hat{\mathbf{x}}_k - \beta\hat{\mathbf{x}}_{k-L})]$. In step 315, the Kalman gain can be determined. The Kalman gain κ_k can be calculated according to

$$\kappa_k = \mathbf{Z}_{k-1}^{-1}\mathbf{B}(\mathbf{I}_{2x2} + \mathbf{D}^T\mathbf{Z}_{k-1}^{-1}\mathbf{B})^{-1}.$$

[0034] In step 320, model outputs can be determined. More particularly, the output estimation y_k can be obtained according to $y_k = \hat{\mathbf{x}}_k^T\mathbf{w}_{k-1}$ and $y_{k-L} = \hat{\mathbf{x}}_{k-L}^T\mathbf{w}_{k-1}$. The method 300 computes the errors at time instant k and $k-L$, where L is the chosen lag.

[0035] In step 325, the error vector $\mathbf{e}_k$ can be determined according to

$$\mathbf{e}_k = \begin{bmatrix} d_k - y_k \\ d_k - y_k - \beta(d_{k-L} - y_{k-L}) \end{bmatrix} = \begin{bmatrix} e_k \\ e_k - \beta e_{k-L} \end{bmatrix}.$$ In step 330 the model of the physical plant, characterized by the weight vector $\mathbf{w}_k$ can be updated according to $\mathbf{w}_k = \mathbf{w}_{k-1} + \kappa_k\mathbf{e}_k$. In step 335, the inverse Hessian $\mathbf{Z}_k^{-1}$ can be updated according to $\mathbf{Z}_k^{-1} = \mathbf{Z}_{k-1}^{-1} - \kappa_k\mathbf{D}^T\mathbf{Z}_{k-1}^{-1}$. The method can loop back to step 310 to repeat as necessary.

[0036] The inventive arrangements presented herein can be used within the context of system identification with noisy inputs. Noisy inputs traditionally lead to biased parameter estimates

that result in poor system identification. The inventive arrangements disclosed herein provide a solution which overcomes the disadvantages of system identification techniques that rely upon optimizing the MSE criterion using techniques such as LMS and Recursive Least Squares (RLS). Techniques such as these do not guarantee unbiased model estimates in noisy conditions. The inventive arrangements disclosed herein can be used to develop accurate model estimates in the presence of noise without adding significant additional computational complexity when compared with MSE-based techniques.

[0037] The various embodiments disclosed herein estimate the optimal EWC solution. The embodiments discussed are derived from the objective function disclosed in equation (3). Once each technique converges, the same weight vector that partially whitens the error signal results.

[0038] The present invention can be realized in hardware, software, or a combination of hardware and software. The present invention can be realized in a centralized fashion in one computer system or in a distributed fashion where different elements are spread across several interconnected computer systems. Any kind of computer system or other apparatus adapted for carrying out the methods described herein is suited. A typical combination of hardware and software can be a general-purpose computer system with a computer program that, when being loaded and executed, controls the computer system such that it carries out the methods described herein.

[0039] The present invention also can be embedded in a computer program product, which comprises all the features enabling the implementation of the methods described herein, and which when loaded in a computer system is able to carry out these methods. Computer program in the present context means any expression, in any language, code or notation, of a set of instructions intended to cause a system having an information processing capability to perform a particular function either directly or after either or both of the following: a) conversion to another language, code or notation; b) reproduction in a different material form.

[0040] This invention can be embodied in other forms without departing from the spirit or essential attributes thereof. Accordingly, reference should be made to the following claims, rather than to the foregoing specification, as indicating the scope of the invention.